NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

DIAGNOSTICS ASSESSMENT PROGRAMME

Diagnostics consultation document

PredictSURE IBD and IBDX to guide treatment of Crohn's disease

The National Institute for Health and Care Excellence (NICE) is producing guidance on using PredictSURE IBD and IBDX in the NHS in England. The diagnostics advisory committee has considered the evidence and the views of clinical and patient experts.

This document has been prepared for public consultation. It summarises the evidence and views that have been considered, and sets out the recommendations made by the committee. NICE invites comments from registered stakeholders, healthcare professionals and the public. This document should be read along with the evidence (the diagnostics assessment report and the diagnostics assessment report addendum).

The advisory committee is interested in receiving comments on the following:

- Has all of the relevant evidence been taken into account?
- Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?
- Are the recommendations sound, and a suitable basis for guidance to the NHS?

Equality issues

NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others. Please let us know if you think that the recommendations may need changing to meet these aims. In particular, please tell us if the recommendations:

- could have a different effect on people protected by the equality legislation than on the wider population, for example by making it more difficult in practice for a specific group to access the technology

Issue date: August 2020

- could have any adverse effect on people with a particular disability or disabilities.

Please provide any relevant information or data you have about such effects and how they could be avoided or reduced.

Note that this document is not NICE's final guidance on PredictSURE IBD and IBDX. The recommendations in section 1 may change after consultation.

After consultation, the committee will meet again to consider the evidence, this document and comments from the consultation. After considering the comments, the committee will prepare its final recommendations, which will be the basis for NICE's guidance on the use of the technology in the NHS in England.

For further details, see the diagnostics assessment programme manual.

Key dates:

Closing date for comments: 28 September 2020

Second diagnostics advisory committee meeting: 7 October 2020

1 Recommendations

1.1 There is not enough evidence to recommend the routine use of the PredictSURE IBD and IBDX tests to help identify people at high risk of severe Crohn's disease and guide treatment.

1.2 Further research is recommended (see section 5) to:

- assess how accurate the tests are for identifying a low or high risk of severe Crohn's disease
- understand how the tests affect decisions about treatment
- assess how the tests affect clinical outcomes.

Why the committee made these recommendations

PredictSURE IBD and IBDX are tests that may be able to identify people at high risk of severe Crohn's disease. If people can be identified in this way, clinicians could

offer the most appropriate treatment to control symptoms while minimising side effects. 'Top-down' treatment has been suggested to be more effective for people with severe Crohn's disease. This reverses the standard order of treatment, starting with biological therapies such as tumour necrosis factor (TNF)-alpha inhibitors.

The clinical evidence for the tests comes from only a small number of people, which does not give confidence in the results reported. There's also not much evidence on how effective top-down treatment is in people who would be classified as high risk using the test, particularly in the NHS where it is not standard practice.

Because of this, the cost-effectiveness estimates are very uncertain and more research is needed. Therefore, these tests are not recommended for routine use in the NHS.

2 The diagnostic tests

Clinical need and practice

2.1 Crohn's disease is a chronic condition that causes inflammation of the gastrointestinal tract, particularly the large intestine and the last section of the small intestine. This condition is characterised by recurring periods of active symptoms (flares). At other times health is generally good (remission). According to Crohn's & Colitis UK, 1 in every 650 people in the UK is affected by Crohn's disease. Complications of Crohn's disease include intestinal strictures (narrowing of the affected area of the intestine), fistulas (ulceration of the lining of the gastrointestinal tract) and perforation. Children may also have growth problems due to poor absorption of nutrients.

2.2 The disease course varies a lot from person to person. Some people are at a higher risk of more frequent flares and relapses that do not respond to standard drug treatment. In the long term, they may be at a higher risk of developing complications and may need surgery.

2.3 Crohn's disease has no cure. The goal of treatment is to induce remission by controlling symptoms and maintain remission to prevent relapse. Current standard care is a 'step-up' strategy, which starts with corticosteroids, then immunosuppressants, then biological therapies if the disease does not respond, or loses response, to treatment. NICE's guideline on Crohn's disease covers the management of Crohn's disease in children, young people and adults and recommends the 'step-up' strategy. Step-up treatment involves multiple courses of steroids before changing to a stronger treatment. 'Accelerated step-up' treatment involves rapidly changing to stronger treatments if the expected response is not seen in the time frame. Adequate response can be defined as no clinical symptoms, no signs of ongoing inflammation, or both.

2.4 An alternative approach not currently recommended as standard care is the 'top-down' strategy, which reverses the order of treatment in the step-up strategy, starting with biologics such as tumour necrosis factor (TNF)-alpha inhibitors. It has been suggested that for some people with Crohn's disease, the top-down strategy could achieve a faster and higher rate of mucosal healing. This could potentially modify the natural disease course and allow people with severe disease to control it better. Biologics are more clinically effective but also associated with more side effects.

2.5 Neither the step-up nor the top-down approach are suitable for all people with Crohn's disease. Clinicians in specialist centres may offer the top-down strategy to people if at diagnosis they consider them to have a poor prognosis (for example, if they have significant fistulising disease, complex perianal disease or multiple risk factors). Being able to predict the course of the disease could enable the clinician to identify people who may benefit from the top-down strategy, that is, early treatment with biologics. There is currently no standard means of categorising people based on their risk of having severe disease.

2.6 PredictSURE IBD and IBDX tests could identify people at a higher risk of severe Crohn's disease, potentially guiding more personalised disease management.

The interventions

PredictSURE IBD

2.7 PredictSURE IBD (PredictImmune) is a whole blood-based biomarker prognostic laboratory-based test combined with a proprietary algorithm to categorise people into a high or low risk of severe Crohn's disease. The test is based on detecting CD8+ T-cell exhaustion. People with a non-exhausted CD8+ T-cell signature were linked to a higher risk of frequently relapsing disease. The test involves isolating mRNA from the whole blood sample using the PAXgene Blood RNA kit (QIAGEN), followed by quantitative polymerase chain reaction with reverse transcription (RT-qPCR) to assess expression of 15 target genes and 2 controls. The RT-qPCR is a 2-step process: cDNA synthesis in the reverse transcription reaction, and then a qPCR on a 384-well plate. A maximum of 4 samples may be analysed per plate because cDNA derived from each RNA sample is run in triplicate.

IBDX

2.8 IBDX (Glycominds) is a panel of 6 indirect solid-phase enzyme-linked immunosorbent assay (ELISA) kits, each of which detects serum levels of specific antiglycan antibodies. Antiglycan antibodies are serological biomarkers thought to be highly specific for Crohn's disease and associated with severe disease. The IBDX ELISA kits available to detect specific antibodies include IBDX anti-chitobioside (ACCA) IgA, IBDX anti-laminaribioside (ALCA) IgG, IBDX anti-mannobioside (AMCA) IgG, IBDX anti-Saccharomyces cerevisiae (gASCA) IgG, IBDX anti-laminarin (anti-L) IgA and IBDX anti-chitin antibody (anti-C) IgA.

The comparator

2.9 The comparator is standard clinical care in which no test or algorithm is used to predict the disease course. Instead, prognosis is based on clinical judgement of presenting signs and symptoms and known clinical risk factors for severe disease.

3 Evidence

The diagnostics advisory committee considered evidence on PredictSURE IBD and IBDX to guide treatment of Crohn's disease from several sources. Full details of all the evidence are in the committee papers.

Clinical effectiveness

3.1 The external assessment group (EAG) systematically reviewed evidence to evaluate the prognostic ability and clinical effectiveness of the PredictSURE IBD and IBDX tests to predict severe disease and guide treatment in people with Crohn's disease who:

- have newly or recently diagnosed disease
- have moderate to severe active disease
- are currently not receiving any concomitant steroids, immunomodulators or biological treatments
- would not have top-down treatment with current standard care in the NHS.

3.2 The EAG identified 8 primary studies (reported in 12 publications) that met the selection criteria for the literature review (see page 18 of the diagnostics assessment report for details of the selection criteria). The studies were all observational. Of the included studies, 7 reported on the diagnostic performance of IBDX. In these studies, a higher number of positive biomarkers was associated with poorer prognosis. Two of the IBDX studies (Wolfel et al. 2017 and Reider et al. 2010c) prospectively assessed the prognostic ability of IBDX for predicting complications (fistulas and stenoses) and Crohn's disease-related surgery. A third

prospective study reported a correlation of IBDX with either a history of complication or surgery at baseline, or their occurrence during follow up (Rieder et al. 2010b). The other 4 studies were cross sectional, reporting a correlation between the number of positive IBDX biomarkers and outcomes associated with severe disease (a presence or history of complications or surgery) at the time of testing. Only 1 study (Biasci et al. 2019) reported on the predictive ability of PredictSURE IBD to classify people into either a high or low risk of severe Crohn's disease, as defined by the study investigators.

3.3 Of the studies identified for IBDX, 3 were carried out in Germany and 1 study each in Canada, France, and the US. One study, published as an abstract, had an unclear location. The study on PredictSURE IBD was carried out across 4 centres in the UK.

Study quality

3.4 The EAG assessed the risk of bias in the included studies using the quality in prognosis studies (QUIPS) tool. All the studies for IBDX were considered to be at a moderate or unclear risk of bias in the measurement of confounding factors domain. Three studies were considered to be at a moderate risk of bias in the participation domain. The study identified for PredictSURE IBD was assessed to be at a low or unclear risk of bias.

Prognostic accuracy

3.5 None of the studies for IBDX was done in only people with newly diagnosed Crohn's disease. The studies had people with an established diagnosis and with a recent diagnosis of Crohn's disease. Median duration of disease at the time of testing ranged from 10.6 months (interquartile range [IQR] 1.7 to 52.3) to 9.4 years (IQR 1 to 44). One prognostic study by Rieder et al. (2010c) assessed the ability of IBDX to predict developing a complication or needing surgery in people with no prior complication or surgery at baseline (n=76). People who tested positive for 2 or more out of 6 IBDX markers had a significantly higher risk

of complications (hazard ratio [HR] 2.5; 95% confidence interval (CI) 1.03 to 6.1; p=0.043), or surgery (HR 3.6; 95% CI 1.2 to 11.0; p=0.023) during the median follow up of 53.7 months, than people who tested positive for 0 or 1 markers. A prognostic study by Wolfel et al. (2017), reported as a conference abstract, showed that the number of positive IBDX markers did not predict a shorter time to repeat intestinal surgery (n=118; median follow up of 100 months).

3.6 Reider et al (2010b) reported that people who had surgery (before or during follow up) had a higher number of positive IBDX markers (median 2.0 [range 1.0 to 3.0]) than those who did not (median 1.0 [range 0.0 to 2.0]; odds ratio [OR] 1.5 [95% CI 1.3 to 1.8]; p<0.001). Similarly, people with a complication had a higher number of positive IBDX markers (median 2.0 [range 1.0 to 3.0]) than those who did not (median 0.0 [range 0.0 to 2.0]; OR 1.5 [95% CI 1.3 to 1.9]; p<0.001). The remaining 4 IBDX studies reported the correlation between IBDX markers and disease phenotype at the time of testing. None of the studies for IBDX estimated sensitivity or specificity.

3.7 Biasci et al. (2019) reported on the prognostic ability of PredictSURE IBD in adults with newly diagnosed Crohn's disease who were not having concomitant treatment. The study included 2 training cohorts (66 people in the biomarker discovery cohort and 39 people in the whole blood classifier cohort) and 1 validation cohort (n=66). The validation cohort and the whole blood classifier cohort were considered relevant to this assessment. In the validation cohort, people categorised as high risk (n=27; 40.9%) had a statistically significantly higher risk of at least 1 treatment escalation than those categorised as low risk (n=39; 59.1%), with a HR of 2.65 (95% CI 1.32 to 5.34; p=0.006). Median duration of follow up was 1.6 years (IQR 1.0 to 3.7) in the high-risk group and 2.4 years (IQR 1.8 to 3.8) in the low-risk group. Sensitivity and specificity for predicting the need for 2 or more escalations in the first 12 months were 77.8% and 70.6% respectively, and within 18 months 72.7% and 73.2%. Negative predictive

value for predicting multiple escalations in the first 18 months was 90.9%. Positive predictive value calculated by the EAG was 42.1%.

Comparative evidence

3.8 A sub-study by Lyons (2020) based on the same cohort as Biasci et al. 2019 and published as an abstract, compared the ability of PredictSURE IBD and IBDX to predict the need for multiple treatment escalations in 74 people with Crohn's disease at Addenbrooke's Hospital, Cambridge. Everyone had active disease at enrolment, and all had accelerated step-up treatment. The author concluded that there was no significant difference between the group who tested positive for at least 1 IBDX marker and those who tested positive for 2 or more IBDX markers, in terms of time to, or frequency of, treatment escalation. In comparison, when the cohort was stratified by PredictSURE IBD, people classed as high risk had a significantly shorter time to treatment escalation than people classed as low risk (p=0.001).

Clinical utility

3.9 No evidence was identified on how the tests affect the decision in clinical practice to offer top-down strategy to people at high risk of severe disease. There was also no evidence on how the tests affect the clinical outcomes of people with severe Crohn's disease.

Cost effectiveness

Systematic review of cost-effectiveness evidence

3.10 The EAG searched for studies on the cost effectiveness of PredictSURE IBD and IBDX in Crohn's disease and economic evaluations of treatments for people with newly diagnosed and moderate to severe Crohn's disease. It did not identify any economic studies for PredictSURE IBD and IBDX, but it did find 11 evaluations relevant to treatment options in Crohn's disease.

3.11 One study by Marchetti et al. (2013) specifically compared the cost effectiveness of top-down (step 1: infliximab plus azathioprine, step 2: additional infliximab plus azathioprine, step 3: methylprednisolone plus azathioprine) and step-up (step 1: methylprednisolone, step 2: methylprednisolone plus azathioprine, step 3: infliximab plus azathioprine) approaches in Italy. The authors concluded that the top-down strategy was better and less costly than the step-up strategy. The treatment strategies modelled in the study by Marchetti are not representative of UK NHS practice. The health economics report for NICE's guideline on Crohn's disease explored the cost effectiveness of 9 induction treatment sequences (composed of 4 treatment lines) for Crohn's disease from the NHS perspective. The remaining 9 studies compared individual treatment steps.

3.12 The company submitted an abstract of a study evaluating the cost effectiveness of PredictSURE IBD to guide early use of biologics in Crohn's disease and ulcerative colitis in the UK. The model structure comprised a decision tree then a Markov transition model. Study results were presented at the European Crohn's and Colitis Organisation Conference in February 2020. The results show that, over a 15-year time horizon, top-down treatment guided by PredictSURE IBD produced an incremental cost-effectiveness ratio (ICER) of £7,179 per quality-adjusted life year (QALY) gained when compared with standard care.

Economic analysis

3.13 The EAG developed a de novo model to assess the cost effectiveness of PredictSURE IBD and IBDX to guide treatment in Crohn's disease. There were no detailed data for IBDX so the EAG assessed its cost effectiveness in an exploratory scenario analysis only.

3.14 The economic analysis was done from the UK NHS and personal social services perspective. The model had a lifetime time horizon (65 years)

with a cycle length of 2 weeks. Costs and benefits were discounted at 3.5% per year.

Model structure

3.15 The model was a hybrid model, with a decision tree for the induction treatment and a Markov transition model for the maintenance treatment. In the induction model, people whose disease does not respond (deterioration; no change; or an improvement of 70 or less in Crohn's Disease Activity Index [CDAI] score) have second-line treatment, according to their treatment allocation (top down or step up). People whose disease responds to the induction treatment (an improvement in CDAI score above 70) move to the maintenance model. They can enter the maintenance model either in remission, mild, or moderate to severe health states. People can then move between these states during maintenance treatment, reflecting the different levels of response to maintenance treatment. People in the mild and moderate to severe states are at risk of relapse and escalating to the next treatment step.

3.16 Escalations from corticosteroids to immunomodulators (step up) and from corticosteroids to tumour necrosis factor (TNF)-alpha inhibitors (top down) were not modelled because in both strategies all people have initial induction treatment with corticosteroids, so they cancel each other out.

3.17 Surgical events are modelled as a standalone outcome in the model, that is, people did not leave their respective health states to enter a surgery health state. Complications and long-term consequences of surgery were not modelled. Time to surgery in the high-risk, top-down cohort was estimated by applying a hazard function generated from the study by Hoekman et al. (2018).

Model inputs

3.18 The population modelled was based on the UK study by Biasci et al. (2019), which the company provided anonymised individual patient data for. There were 105 people in the cohort with Crohn’s disease; 88 were

newly diagnosed. However, the EAG based its analysis on 40 people in the study whose treatment matched the standard definition of step-up treatment in the UK, that is, people who had first-line treatment with corticosteroids and second-line treatment with immunomodulators (after failure of corticosteroids). This informed the estimates of time to treatment escalation and time to surgery in the base case. To extrapolate time to treatment escalation data to the time horizon of the model, the EAG used individual patient data to generate time to event data for time to first escalation.

3.19 D'Haens et al. (2008) and its 10-year follow-up study by Hoekman et al. (2018) informed estimates for effectiveness of top-down compared with step-up treatment. The study by D'Haens was a 2-year multicentre randomised trial that assessed the clinical efficacy of early combined immunosuppression (top-down treatment) compared with conventional treatment (step-up treatment) in people with newly diagnosed Crohn's disease. People randomised to top-down treatment had induction treatment with infliximab and azathioprine. People had no infliximab maintenance but were allowed infliximab as needed and, if necessary, corticosteroids, to control disease activity. People randomised to step-up treatment had corticosteroids, followed, in sequence, by azathioprine and infliximab. The study by Hoekman et al. (2018) retrospectively reviewed the medical records of people included in the D'Haens trial, which collected data on hospitalisation, flares, surgery, clinical activity and other outcomes, for a median follow up of 10 years.

Effectiveness of induction and maintenance therapies

3.20 Probabilities of response and remission with induction and maintenance therapies were based on data from a pragmatic search and from NICE's guidance on vedolizumab for treating moderately to severely active Crohn's disease after prior therapy. Based on this guidance, the EAG estimated that 21.2% of responders remained in the moderate to severe

disease state. The probability of response is the same for top down and step up, except for immunomodulators in the step-up strategy (table 1).

Table 1 Probability of response and remission with induction and maintenance therapies

Treatment strategy	Induction: response	Induction: remission	Maintenance: response	Maintenance: remission
Top down: biologics	32%	13%	2%	28%
Top down: anti-TNF	26%	37%	10%	33%
Step up: biologics	32%	13%	2%	28%
Step up: anti-TNF	26%	37%	10%	33%
Step up: immunomodulator	23%	16%	15%	25%

3.21 The costs considered in the model are the costs of the diagnostic tests, treatment and care of Crohn's disease. The total cost of testing charged by the laboratory was £1,250 for PredictSURE IBD and £347 (estimated) for IBDX. Table 2 shows the dose prices and induction dosages for induction treatment in top-down and step-up strategies, taken from BNF and NHS reference costs, and maintenance treatment dosages based on clinical opinion.

Table 2 Treatment doses and costs for induction and maintenance therapies

Treatment	Dose per unit (mg)	List price per unit	Induction dosages	Maintenance dosages
Ustekinumab	130	£2,147.00	Induction dose at week 0 depends on body weight: 260 mg for 56 kg 390 mg for 56 kg to 85 kg 520 mg for 86 kg or over	90 mg every 8 weeks
Vedolizumab	300	£2,050.00	300 mg at week 0, 2 and 6	300 mg every 8 weeks
Infliximab	100	£377,66	5 mg/kg at week 0, 2 and 6	5 mg/kg every 8 weeks
Adalimumab	40	£308.13	160 mg at week 0; 80 mg at week 2	40 mg every 2 weeks
Azathioprine	50	£0.04	2.5 mg/kg/week for 8 weeks	2.5 mg/kg/week
6-MP	50	£1.97	1.25 mg/kg/week	1.25 mg/kg/week
Methotrexate	25/15	£16.64 or £14.92	25 mg/week for 8 weeks	15 mg/week
Prednisolone	2.5	£0.04	40 mg; tapered by 5 mg per week – 8 weeks total	No maintenance with prednisolone
Intravenous administration (outpatient)	1	First: £199 Follow up: £212	Not applicable	Not applicable

The total cost of managing maintenance health states for 2 weeks was £17 for remission, £27 for a mild state and £122 for a moderate to severe state. This included outpatient, radiology, endoscopy and hospitalisation costs.

3.22 The EAG used the utility values from NICE's guidance on vedolizumab (based on EQ-5D data from GEMINI studies) in the base case analysis and a mapping algorithm based on NICE's guidance on ustekinumab for moderately to severely active Crohn's disease after prior therapy in a scenario analysis. All utilities were adjusted to account for the age and sex of the modelled population, according to Ara and Brazier 2010. Surgery-related disutility was estimated from the Marchetti study. Table 3 shows the utility values used in the modelling.

Table 3 Utility values used for remission, mild, moderate to severe health states

Health state	NICE guidance on vedolizumab	NICE guidance on ustekinumab
Remission	0.820	0.820
Mild disease	0.730	0.700
Moderate to severe	0.570	0.550

Key assumptions

3.23 The EAG assumed that:

- PredictSURE IBD (and IBDX in the scenario analysis) are 100% accurate in categorising people into high and low risk of severe disease.
- People categorised as high risk by the test have top-down treatment.
- People have the same baseline probability of escalating to the next step in the step-up strategy (estimated from time to first escalation in Biasci et al.) regardless of the number of previous escalations.
- 30% of people receiving anti-TNF and 20% of people receiving non-anti-TNF biologics have combination treatment with immunomodulators.
- Response to anti-TNF does not depend on the prior lines of treatment.
- People in the top-down strategy have a longer time to treatment escalation and a longer time to surgery than people in the step-up strategy, based on extrapolation of results from D'Haens et al. (2008) and Hoekman et al. (2018).

3.24 To use the D'Haens study the EAG assumed that:

- The relative treatment effect of top-down and step-up strategies in a mixed-risk population is the same as the relative treatment effect in a high-risk population.

- Time to relapse is a proxy measure for time to the next treatment escalation.
- The effectiveness of treatment strategies in this study is a proxy for the treatment effectiveness of the first step in the top-down (anti-TNF) and step-up (immunomodulators) strategies modelled.

3.25 To estimate the relative treatment effect of top-down and step-up treatment on time to treatment escalation, the EAG digitised the time to relapse Kaplan–Meier data from D'Haens et al. to estimate a hazard function. This was applied to the first treatment step in the high-risk, top-down arm of the model.

3.26 The base case was revised to reflect the assumption that time to treatment escalation restarts on each new treatment rather than reducing over time and as treatment sequences progress. Cost-effectiveness results presented are from the revised base case.

Base case results

3.27 Results of the revised base case (detailed in the addendum to the diagnostics assessment report) superseded results of the primary analysis. In the revised analysis the time to treatment escalation restarts on each new treatment.

3.28 The base case compared the top-down strategy (using PredictSURE IBD to predict who was high risk) with standard care, in which a high-risk person has step-up treatment. In both the deterministic and probabilistic analyses PredictSURE IBD was dominated by standard care, meaning it costs more and has less QALYs.

- Deterministic result: incremental cost was £9,084 and incremental QALY was -0.08.
- Probabilistic result: incremental cost was £12,132 and incremental QALY was -0.03.

The testing strategy had a less than 10% probability of being cost effective against standard care at the maximum acceptable ICERs of £20,000 and £30,000 per QALY gained.

3.29 The deterministic fully incremental cost-effectiveness analysis explored as a scenario analysis showed that PredictSURE IBD (incremental cost £903, incremental QALY 0) and IBDX (incremental cost £8,181, incremental QALY -0.08) were dominated when compared with standard care of no testing. In the absence of robust evidence on the prognostic accuracy of both tools, the cost-effectiveness analysis only differs in the cost of the tests.

Cost-effectiveness results: scenario analyses

3.30 The dominance of the step-up strategy may be because of the benefit some people get from having immunomodulators first, before biologics. The clinical experts told the EAG that people on the top-down strategy do not have immunomodulators after 3 lines of biologics. However, the EAG explored a scenario which had immunomodulators as the last treatment option in the top-down arm. Deterministic base case results for this scenario showed that PredictSURE IBD (top-down strategy) generated 0.07 more QALYs than the step-up strategy, at an additional cost of £7,502, producing an ICER of £105,148 per QALY gained. This is higher than £20,000 to £30,000 per QALY gained, the range NICE normally considers an acceptable use of NHS resources.

3.31 The EAG ran a series of individual scenario analyses, most of which showed that PredictSURE IBD was dominated by standard care. The EAG also ran a combination of individual scenarios, because these were thought to have more impact than individual scenarios.

3.32 If the analysis assumed that the condition did not respond to treatment with immunomodulators for any high-risk person in the step-up arm, so they had no benefit from them, PredictSURE IBD had an ICER of £170,180 per QALY gained. The proportion of people who responded to

immunomodulators was then varied. This showed that the 2 strategies became clinically equivalent when it was assumed that 97% of high-risk people in the step-up arm do not benefit from immunomodulators.

3.33 The EAG explored a scenario that assumed PredictSURE IBD had a lower test accuracy, and the effect of misdiagnosis. In this scenario PredictSURE was more costly and generated a QALY gain of 0.15, producing an ICER of £64,876 per QALY gained. This gain in QALY, despite the lower accuracy of the test, can be attributed to the assumption that some lower-risk people misdiagnosed as high risk go on to have top-down treatment, without the need for further escalation.

3.34 Assumptions about treatment discontinuation were based on Marchetti 2013, which reported that 76% of people had mucosal healing after 2 years in remission with biologic treatments using a top-down strategy and 40% using a step-up strategy. In a scenario analysis that assumed 76% of people in the top-down arm and 40% of people in the step-up arm discontinued biologics, PredictSURE IBD was less costly and less effective than standard care, producing an ICER of £46,263 per QALY gained. Scenarios combining the effect of misdiagnosis with the same proportion of people discontinuing biologics in both arms, that is, 40% in step up and top down or 76% in step up and top down, produced ICERs of £48,034 and £32,875 per QALY gained respectively.

3.35 Individual scenarios were combined to explore the impact of increasing the effectiveness of the top-down strategy while reducing the treatment cost of biologics. The results of these combined scenarios varied. One scenario combined 3 assumptions, that:

- base case risk of relapse for second and later treatment steps is the same
- discontinuation of biologic treatment is 76% for top down and step up
- 100% of people in the step-up arm do not respond to immunomodulators.

This produced an ICER of £29,225 per QALY gained in favour of top down.

3.36 A tornado plot of the one-way sensitivity analyses showed that response to biologics in the top-down arm of the model was a key driver of the deterministic ICER.

4 Committee discussion

Clinical effectiveness

Knowing the likely course of the disease may help people with Crohn's disease and the NHS

4.1 The patient expert explained that having Crohn's disease can substantially affect the quality of life of the person and their family. Currently the extent of inflammation is monitored using endoscopic imaging and faecal calprotectin blood tests, but they do not predict disease progression or the likelihood of needing surgery in the future. People may not want invasive monitoring using colonoscopy because it is stressful to prepare for, has unpleasant side effects and may aggravate symptoms. The patient expert suggested that a test to predict long-term disease course could help give people a better understanding and acceptance of their condition, and make planning review appointments more efficient.

Studies on the prognostic ability of the tests are heterogenous and have small sample sizes

4.2 The reviewed studies on prognostic ability had mixed populations, including people with ulcerative colitis. The numbers of people with Crohn's disease in each study was small, given the prevalence of the condition in the wider population. The committee noted that the small sample sizes could mean that the reviewed studies were underpowered to

produce robust estimates of the prognostic ability of the tests. The committee also noted that there are other predictive studies for Crohn's disease with larger populations, showing that larger sample sizes are possible. The committee concluded that the heterogeneity in the population and the population size added substantial uncertainty to the interpretation of study results.

There is no standard definition of a high or low risk of severe disease

4.3 The reviewed studies used different measures to define a person as being at high or low risk of severe Crohn's disease. IBDX studies used poor outcomes, such as surgery and complications, as a proxy for severe disease (see sections 3.5 and 3.6), whereas the PredictSURE IBD study used the need for multiple treatment escalations (see section 3.7). This inconsistency is a source of additional uncertainty.

The accuracy of PredictSURE IBD and IBDX in predicting severe disease is uncertain

4.4 Little data were identified on the prognostic accuracy of the tests. Sensitivity, specificity and negative predictive value were only reported for the PredictSURE IBD test, and in only 1 study (Biasci 2019). The clinical expert said that at the moment severe disease may be predicted by known risk factors such as age and smoking status. But there is no consensus on, or algorithm for, how these risk factors should be combined, and their predictive value is limited. The clinical expert also said that, based on the findings of the Biasci study, the PredictSURE IBD test appears to perform better than risk prediction based on clinical features or endoscopic findings, and therefore has the potential to be a useful test. The committee noted that it would help to understand if the tests can give a more accurate prognosis when used alongside clinical features rather than as a substitute. The committee concluded that overall, the evidence on the prognostic accuracy of PredictSURE IBD and IBDX is weak, and encouraged further research on their accuracy used alongside clinical features (see section 5.1).

There is little evidence on how the tests affect treatment decisions

4.5 The proposed value of the tests is to categorise people with Crohn's disease according to their risk of severe disease. People predicted to have severe disease could have top-down treatment, which may help control the disease early, leading to better outcomes like fewer flare-ups, and prevent bowel damage and limit the need for surgery. The committee noted that currently there was no evidence on how the tests can help with decisions about personalised treatment plans. It concluded that it would help to have research on how the tests affect treatment decisions (see section 5). PROFILE, a randomised, multicentre, biomarker-stratified, open-label study is ongoing in the UK with results expected in 2022. This trial uses PredictSURE IBD to assign people to top-down or step-up treatment, and may help address this evidence gap.

There is no evidence on how the tests affect clinical outcomes

4.6 The committee considered that there was no evidence to show that using the prognostic tests to identify people at high risk of severe disease and help guide treatment improves clinical outcomes. The committee encouraged studies assessing how the tests affect clinical outcomes (see section 5).

Cost effectiveness

Drug treatment for people with Crohn's disease varies across the NHS

4.7 The committee noted that the treatment sequences modelled by the external assessment group (EAG) may not reflect treatment in the NHS. The EAG said that in its model 30% of people who had a tumour necrosis factor (TNF)-alpha inhibitor, and 20% of people who had a biological treatment that was not an anti-TNF, also had an immunomodulator. This is because there is evidence to show that combination treatment reduces the chances of losing response to biologics (immunogenicity). However, clinical experts said there is no consensus on using monotherapy or combination therapy, and that it varies in clinical practice. There is also

the option of having immunomodulators after biologics as part of treatment de-escalation in the top-down strategy (as modelled by the company – see section 4.12). Top-down treatment is not widely used in the NHS and so it is uncertain what the treatment pathway would look like. The company model included an immunomodulator step after biologics but the EAG base case did not. The EAG explored this as a scenario analysis (see section 3.30). In addition, the biologics modelled as second and third line can also be used as first line. The committee concluded that variation in clinical practice created an added level of uncertainty around the model structure.

It's not certain if top-down treatment has clinical benefits over step-up treatment

4.8 The committee heard from clinical experts that early rather than late treatment with biologics could improve outcomes for people likely to have more severe disease. The EAG noted that the evidence on the effectiveness of top-down compared with step-up treatment in the model was from the D'Haens study. This showed that people who had top-down treatment had a longer time to relapse than people who had step-up treatment. The hazard function (based on the assumption that time to relapse is a proxy for time to next treatment escalation) derived from D'Haens was applied only to the first step of the model (the anti-TNF compared with immunomodulator step). Later treatment steps in both the top-down and step-up strategies were assumed to have the same time to treatment escalation as anti-TNF in the top-down arm. This assumption was made because there was no evidence either way. The top-down treatment sequence modelled in D'Haens differed from the one described by the clinical experts because people did not carry on having maintenance treatment with infliximab but were allowed infliximab as needed (see section 3.19). This might have underestimated the benefits of top-down treatment. The EAG said that in the long term top down may not have an advantage over step up because the 10-year follow-up study

of D'Haens (Hoekman 2018) showed no difference in hospitalisation, surgery and endoscopic remission between both strategies. The clinical experts considered that early treatment with biologics does make a difference, but good-quality evidence generalisable to the NHS to support this is limited. Registry data could have been useful. The committee concluded that more evidence is needed on the effectiveness of top-down compared with step-up strategies. This is because if there is no evidence of benefit, there is no clinical rationale for identifying people at high risk of severe disease and treating them using a top-down strategy.

Because of the lack of data and the need for many assumptions, the model results are not certain

4.9 The committee noted that interpreting the modelling was difficult because of the very weak data feeding into it. There were limited data on the prognostic accuracy of the tests (see section 4.4), on the effectiveness of a top-down strategy compared with a step-up strategy (see section 4.8), and no information from studies on how these 2 steps would combine to affect clinical outcomes. The committee heard that the EAG had to make many assumptions to be able to link the evidence in the model. There was great variation in the results of the model. Base case results (see section 3.28) showed that standard care dominated the top-down strategy. This dominance was sustained in the majority of the scenario analyses, and the probabilistic sensitivity analysis scatter plot showed the top-down strategy was mostly more costly and less effective than standard care. Some of the scenario analyses produced incremental cost-effectiveness ratios (ICERs) in favour of the top-down strategy, although these were far higher than the range that NICE usually considers to be cost effective (see sections 3.30 to 3.34). One scenario (see section 3.35) combining multiple assumptions produced an ICER in the acceptable range in favour of the top-down strategy. Because of the limited data and assumptions that needed to be made, the cost effectiveness of the tests is highly uncertain.

Assuming that IBDX and PredictSURE IBD have the same prognostic ability is not appropriate

4.10 Only data on the prognostic ability of PredictSURE IBD were included in the base case. The EAG included IBDX in an exploratory analysis that assumed that the ability of IBDX to identify people at high or low risk was the same as PredictSURE IBD. The committee heard that the tools identify different markers and require different test samples. The committee also noted that there was 1 abstract (Lyons 2020), which compared both tools and showed that PredictSURE IBD predicted a shorter time to treatment escalation in people classed as high risk. IBDX did not predict a difference in time to treatment escalation between people positive for 2 or more markers and those positive for only 1 marker (see section 3.8). The committee concluded it was not appropriate to assume the tests had the same prognostic accuracy, and that more evidence is needed (see section 5).

Some of the key assumptions in the model are drivers of the model results

4.11 The model results were mainly driven by the assumption that step-up treatment has benefits over top-down treatment because of the proportion of people who respond to immunomodulators in the step-up arm. The EAG noted that having the immunomodulator step at the start of the step-up strategy meant that, for some high-risk people, their condition could respond to less costly immunomodulators. Another assumption that drove the model results was that after 2 years in remission with biologics, a proportion of people have mucosal healing and do not need more treatment escalations. A scenario in which some low-risk people were assumed to be misdiagnosed as high risk (see section 3.33) showed QALYs being gained in favour of PredictSURE IBD because they did not need any more treatment escalation.

The EAG's model results are different from the company's model and the most relevant published economic model

4.12 The base case probabilistic and deterministic results of the EAG's model produced QALYs in favour of standard care. This suggests that a no testing strategy with step-up treatment is better for people at high risk of severe Crohn's disease than top-down treatment using the prognostic tool. This result was not consistent with the company's model and the model reported by Marchetti (2013), both of which reported that a top-down strategy is associated with more QALYs. The EAG noted that the difference between its model and the company's was that the treatment sequence modelled by the company had an immunomodulator as a last treatment step in the top-down arm. This was not modelled in the EAG's base case but as a scenario analysis. This scenario produced an ICER in favour of top-down treatment that was much higher than what NICE normally considers a cost-effective use of NHS resources (see section 3.30). The company's model also assumed a constant relative treatment effect, whereas the EAG's model assumed a diminishing relative treatment effect (see further details in the addendum to the diagnostic assessment report). Marchetti modelled a different treatment sequence (see section 3.11) to the EAG's, and a different time horizon – 5 years compared with the EAG's 65 years. The EAG did not explore changing the time horizon so it was not clear if the time horizon influenced the different results. The difference in the results was likely due to the uncertainties in the top-down treatment pathway and the effectiveness of top-down compared with step-up strategies.

Evidence from a different starting cohort that includes children and teenagers would be useful

4.13 The committee heard that the average age in the EAG's model was 35. It considered that the model might not reflect other age groups that are first diagnosed with Crohn's, for example, one peak is in teenagers and another is at around 60. A clinical expert noted that the treatment pathway

for children or teenagers would be different from adults because children often follow a more severe disease course and may need enteral nutrition. The committee heard that modelling this population may require an entirely new model rather than an adaptation of the model built by the EAG for the adult population.

Modelling adverse events or varying the cost of surgery may not have a huge impact on the results

4.14 The EAG did not model adverse events, to keep the model simple. It predicted that if it had modelled adverse events top-down treatment would have been more dominated. The committee thought the cost of surgery might have been underestimated and that its impact on the model results was not clear. The EAG noted that, although it did not vary the costs of surgery, the number of surgical events modelled was very small, so it did not anticipate a significant difference in results.

Multiple uncertainties make it difficult to determine cost effectiveness so the tests cannot be recommended for routine use in the NHS

4.15 Lack of evidence on the prognostic ability, the effect on treatment decisions and clinical outcomes (see sections 4.4 to 4.6) of the PredictSURE IBD and IBDX tests makes it difficult to assess the cost effectiveness of the tests for assigning people to top-down or step-up treatment. The base case model was based on data for PredictSURE IBD. IBDX was only included in an exploratory scenario analysis (see section 3.13). Issues in the modelling (see sections 4.7 to 4.10) relate to:

- the effectiveness of the top-down strategy
- the assumed equivalence in prognostic accuracy of both tools
- how appropriate the sequence modelled is to all people with Crohn's disease.

These, and the many assumptions needed to link the data because of limited evidence, make the cost effectiveness of the tests to the NHS

uncertain. In the absence of the evidence the committee would have liked to see (see section 5), changes to the model at this time would not change the overall conclusion.

5 Recommendations for further research

5.1 The committee recommended more research on:

- the accuracy of PredictSURE IBD and IBDX in identifying people at high or low risk of severe Crohn's disease
- how PredictSURE IBD and IBDX, when used alongside clinical features, affect clinical decisions about whether step-up or top-down treatment is offered
- how PredictSURE IBD and IBDX affect clinical outcomes once someone has been assigned to top-down or step-up treatment, taking into account the different pathways that children and adults may follow.

6 Implementation

NICE intends to develop tools, in association with relevant stakeholders, to help organisations put this guidance into practice.

In addition NICE will support this guidance through a range of activities to promote the recommendations for further research. The research proposed will be considered by the NICE Medical Technologies Evaluation Programme research facilitation team for developing specific research study protocols as appropriate. NICE will also incorporate the research recommendations in section 5 into its guidance research recommendations database and highlight these recommendations to public research bodies.

7 Review

NICE reviews the evidence 3 years after publication to ensure that any relevant new evidence is identified. However, NICE may review and update the guidance at any time if significant new evidence becomes available.

Mark Kroese

Chair, diagnostics advisory committee

August 2020

8 Diagnostics advisory committee members and NICE project team

Committee members

This topic was considered by the diagnostics advisory committee, which is a standing advisory committee of NICE.

Committee members are asked to declare any interests in the test to be assessed. If it is considered there is a conflict of interest, the member is excluded from participating further in that assessment.

The minutes of each committee meeting, which include the names of the members who attended and their declarations of interests, are posted on the NICE website.

Additional specialist committee members took part in the discussions for this topic:

Specialist committee members

Professor Shaji Sebastian

Consultant Gastroenterologist, Hull Teaching Hospitals NHS Trust

Professor Jack Satsangi

Professor of Gastroenterology, University of Oxford

Dr Jenny Epstein

Consultant Paediatric Gastroenterologist, Chelsea and Westminster Hospital

Tracey Tyrrell

IBS Advanced Nurse Practitioner, London North West University Healthcare Trust

Rebecca Harmston

Lay member

Zehra Arkir

Consultant clinical scientist (biochemist) biochemical sciences, Viapath

NICE project team

Each diagnostics assessment is assigned to a team consisting of a technical analyst (who acts as the topic lead), a technical adviser and a project manager.

Ewa Rupniewska

Topic lead (until December 2019)

Tosin Oladapo

Topic lead (from February 2020)

Frances Nixon

Technical adviser

Donna Barnes

Project manager

ISBN:

Issue date: August 2020